(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 1 391 226 A1

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
25.02.2004 Bulletin 2004/09

(51) Int Cl.[7]: **A63F 13/12**, H04N 7/24, H04N 7/173

(21) Application number: **02360239.4**

(22) Date of filing: **12.08.2002**

(84) Designated Contracting States:
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR IE IT LI LU MC NL PT SE SK TR
Designated Extension States:
AL LT LV MK RO SI

(71) Applicant: **ALCATEL**
75008 Paris (FR)

(72) Inventor: **Domschitz, Peter**
70191 Stuttgart (DE)

(74) Representative: **Brose, Gerhard et al**
Alcatel,
Intellectual Property Department Stuttgart
70430 Stuttgart (DE)

(54) Method and devices for implementing highly interactive entertainment services using interactive media-streaming technology, enabling remote provisioning of virtual reality services

(57) The invention relates to a method for generating an interactive virtual reality with a network service using interactive media-streaming technology comprising the steps of establishing an action stream session comprising connection handling, quality of service handling, adapting the network environment by demanding network resources and control information, establishing media-streaming path from the service to the client and a user interaction control path in the reverse direction, controlling the network with respect to required quality of service, continuously, generating and transmitting individual media streams to the client (ASC) by embedding interaction into a virtual reality, and extracting and encoding a media stream at the service using a virtual reality description compressed motion picture stream, encoding and transmitting the user's interaction to the service, as well as de-coding and playing the individual media data stream at the client side. Further it relates to a Action Streaming Service, Action Streaming Client (ASC), Action Streaming Server (ASS), Action Streaming System, Action Stream, Action Streaming Session, Action Streaming Protocol, and Computer Software Products for generating an interactive virtual reality.

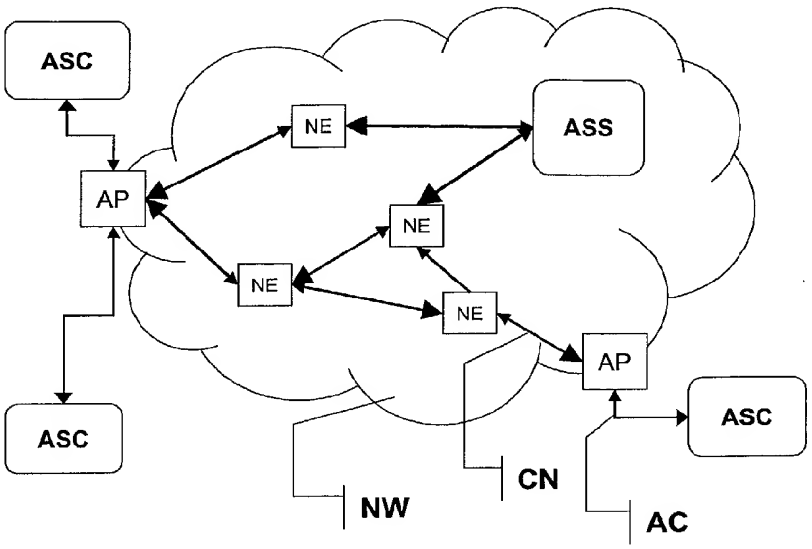


Figure 3.

EP 1 391 226 A1

Description

BACKGROUND OF THE INVENTION

Field of the Invention

[0001] The present invention relates to the provisioning of highly interactive video/audio services, e.g. remote gaming, with reactive requirements and hard real-time conditions on required reactive and realistic dynamic visualization. More particularly, the present invention relates to a method, an action streaming service, an action streaming client, an action streaming server, an action streaming system, an action stream, an action streaming session, an action streaming protocol, and computer software products for generating an interactive virtual reality.

Background

[0002] The real-time video/audio processing for electronic gaming and other interactive virtual reality based entertainment requires specialized and performant local devices like high-end personal computers or game consoles.

[0003] There are multiple games for personal computers and consoles available allowing a plurality of player participating a (shared) game. The devices uses access network technology to share a virtual world. This is done using e.g. the Internet to exchange and align the virtual worlds. To minimize the consumed network resource a common used technique is to parameterize such virtual world.

[0004] For instance the virtual world of a soccer game is identified by the playing team and the location. The visualization of the location, i.e. the playground might be a part of the local game software itself. Hence the short string "WORLD CUP 2002 FINAL" specifies completely the players and the playground graphics. The state of the game could be specified by the orientation and position of the players and the ball. The classical distributed game architecture is to align these states via a network, e.g. the Internet, and generating the virtual reality, meaning the video and audio, locally at a game console comprising perspectives, models, and rendering. This approach avoids heavily interchanging data across the network.

[0005] The above architecture was influenced by missing network resources, namely bandwidth or delay. In the future the situation becoming slightly different. Digital video and audio is an emerging technology, deploying digital encoded audio and video streams. To support this kind of network applications the European Telecommunications Standards Institute (ETSI) designed a standard platform, the Media Home Platform.

Media Home Platform

[0006] The Multimedia Home Platform (MHP) defines a generic interface between interactive digital applications and the terminals on which those applications execute. This interface decouples different provider's applications from the specific hardware and software details of different MHP terminal implementations. It enables digital content providers to address all types of terminals ranging from low-end to high-end set top boxes, integrated digital TV sets and multimedia PCs. The MHP extends the existing, successful Digital Video Broadcast (DVB) standards for broadcast and interactive services in all transmission networks including satellite, cable, terrestrial, and microwave.

[0007] The architecture of the MHP is defined in terms of three layers: resources, system software and applications. Typical MHP resources are MPEG processing, I/O devices, CPU, memory and a graphics system. The system software uses the available resources in order to provide an abstract view of the platform to the applications. Implementations include an application manager (also known as a "navigator") to control the MHP and the applications running on it.

[0008] The core of the MHP is based around a platform known as DVB-J. This includes a virtual machine as defined in the Java Virtual Machine specification from Sun Microsystems. A number of software packages provide generic application program interfaces (APIs) to a wide range of features of the platform. MHP applications access the platform only via these specified APIs. MHP implementations are required to perform a mapping between these specified APIs and the underlying resources and system software.

[0009] The main elements of the MHP specification are:

- MHP architecture (as introduced above),
- definition of enhanced broadcasting and interactive broadcasting profiles,
- content formats including PNG, JPEG, MPEG-2 Video/Audio, subtitles and resident and downloadable fonts,
- mandatory transport protocols including DSM-CC object carousel (broadcast) and IP (return channel),
- DVB-J application model and signaling,
- hooks for HTML content formats (DVB-HTML application model and signaling),
- DVB-J platform with DVB defined APIs and selected parts from existing Java APIs, JavaTV, HAVi (user interface) and DAVIC APIs,
- security framework for broadcast application or data authentication (signatures, certificates) and return channel encryption (TLS),
- graphics reference model.

[0010] The MHP specification provides a consistent set of features and functions required for the enhanced

broadcasting and interactive broadcasting profiles. The enhanced broadcasting profile is intended for broadcast (one way) services, while the interactive broadcasting profile supports in addition interactive services and allows MHP to use the world-wide communication network provided by the Internet.

[0011] The MHP therefore is simply a common Application Program Interface (API) that is completely independent of the hardware platform it is running on. Enhanced Broadcasts, Interactive Broadcasts and Internet Content from different providers can be accessed through a single device e.g. Set top box or IDTV, that uses this Common DVB-MHP API.

It will enable a truly horizontal market in the content, applications and services environment over multiple delivery mechanisms (Cable, Satellite, Terrestrial, etc.).

Encoding Audio and Video Streams

[0012] Crucial for deploying interactive audio/video-streaming is encoding and decoding. In this area MPEG (pronounced M-peg), which stands for Moving Picture Experts Group, is the name of family of standards used for coding audio-visual information, e.g. movies, video, music in a digital compressed format. MPEG uses very sophisticated compression techniques.

[0013] MPEG-1 is a coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s. It addresses the problem of combining one or more data streams from the video and audio parts of the MPEG-1 standard with timing information to form a single stream. This is an important function because, once combined into a single stream, the data are in a form well suited to digital storage or transmission.

[0014] It specifies a coded representation that can be used for compressing video sequences - both 625-line and 525-lines - to bit-rates around 1,5 Mbit/s. It was developed to operate principally from storage media offering a continuous transfer rate of about 1,5 Mbit/s. Nevertheless it can be used more widely than this because the approach taken is generic.

[0015] A number of techniques are used to achieve a high compression ratio. The first is to select an appropriate spatial resolution for the information. The algorithm then uses block-based motion compensation to reduce the temporal redundancy. Motion compensation is used for causal prediction of the current picture from a previous picture, for non-causal prediction of the current picture from a future picture, or for interpolative prediction from past and future pictures. The difference signal, the prediction error, is further compressed using the discrete cosine transform (DCT) to remove spatial correlation and is then quantised. Finally, the motion vectors are combined with the DCT information, and coded using variable length codes.

[0016] MPEG-1 specifies a coded representation that can be used for compressing audio sequences - both mono and stereo. Input audio samples are fed into the encoder. The mapping creates a filtered and sub-sampled representation of the input audio stream. A psychoacoustic model creates a set of data to control the quantiser and coding. The quantiser and coding block creates a set of coding symbols from the mapped input samples. The block 'frame packing' assembles the actual bit-stream from the output data of the other blocks, and adds other information, e.g. error correction if necessary.

[0017] MPEG-2 describes a generic coding of moving pictures and associated audio information addresses the combining of one or more elementary streams of video and audio, as well as, other data into single or multiple streams which are suitable for storage or transmission. This is specified in two forms: the Program Stream and the Transport Stream. Each is optimized for a different set of applications. The Program Stream is similar to MPEG-1 Systems Multiplex. It results from combining one or more Packetized Elementary Streams (PES), which have a common time base, into a single stream. The Program Stream is designed for use in relatively error-free environments and is suitable for applications which may involve software processing. Program stream packets may be of variable and relatively great length.

[0018] The Transport Stream combines one or more Packetized Elementary Streams (PES) with one or more independent time bases into a single stream. Elementary streams sharing a common time-base form a program. The Transport Stream is designed for use in environments where errors are likely, such as storage or transmission in lossy or noisy media.

[0019] MPEG-2 builds on the powerful video compression capabilities of MPEG-1 to offer a wide range of coding tools. These have been grouped in profiles to offer different functionalities.

[0020] MPEG-2 Digital Storage Media Command and Control (DSM-CC) is the specification of a set of protocols which provides the control functions and operations specific to managing MPEG-1 and MPEG-2 bit-streams. These protocols may be used to support applications in both stand-alone and heterogeneous network environments. In the DSM-CC model, a stream is sourced by a Server and delivered to a Client. Both the Server and the Client are considered to be Users of the DSM-CC network. DSM-CC defines a logical entity called the Session and Resource Manager (SRM) which provides a (logically) centralized management of the DSM-CC Sessions and Resources.

[0021] MPEG-4 builds on the three fields: Digital television, Interactive graphics applications (synthetic content), and Interactive multimedia (World Wide Web, distribution of and access to content). MPEG-4 provides the standardized technological elements enabling the integration of the production, distribution and content access paradigms of the three fields. The following sections illustrate the MPEG-4 functionalities described above, using the audiovisual scene depicted in Figure 2.

Coded representation of media objects

[0022] MPEG-4 audiovisual scenes are composed of several media objects, organized in a hierarchical fashion. At the leaves of the hierarchy, one find primitive media objects, such as:

- Still images, e.g. as a fixed background,
- Video objects, e.g. a talking person - without the background,
- Audio objects, e.g. the voice associated with that person, background music. MPEG-4 provides a number of such primitive media objects, capable of representing both natural and synthetic content types, which can be either 2- or 3-dimensional. In addition to the media objects mentioned above and shown in Figure 1, MPEG-4 defines the coded representation of objects such as text and graphics, talking synthetic heads and associated text used to synthesize the speech and animate the head; animated bodies to go with the faces, or synthetic sound.

[0023] A media object in its coded form consists of descriptive elements that allow handling the object in an audiovisual scene as well as of associated streaming data, if needed. It is important to note that in its coded form, each media object can be represented independent of its surroundings or background.

[0024] The coded representation of media objects is as efficient as possible while taking into account the desired functionalities. Examples of such functionalities are error robustness, easy extraction and editing of an object, or having an object available in a scaleable form.

Composition of media objects

[0025] Figure 2 explains the way in which an audiovisual scene in MPEG-4 is described as composed of individual objects. The figure contains compound media objects that group primitive media objects together. Primitive media objects correspond to leaves in the descriptive tree while compound media objects encompass entire sub-trees. As an example: the visual object corresponding to the talking person and the corresponding voice are tied together to form a new compound media object, containing both the aural and visual components of that talking person. Such grouping allows authors to construct complex scenes, and enables consumers to manipulate meaningful (sets of) objects.

[0026] More generally, MPEG-4 provides a way to describe a scene, allowing for example to:

- place media objects anywhere in a given coordinate system,
- apply transforms to change the geometrical or acoustical appearance of an object,
- group primitive media objects in order to form compound media objects;
- apply streamed data to media objects, in order to modify their attributes (e.g. a sound or animation parameters driving a synthetic face);
- change, interactively, the user's viewing and listening points anywhere in the scene.

[0027] The scene description builds on several concepts from the Virtual Reality Modeling language (VRML) in terms of both its structure and the functionality of object composition nodes.

Description and synchronization of streaming data for media objects

[0028] Media objects may need streaming data, which is conveyed in one or more elementary streams. An object descriptor identifies all streams associated to one media object. This allows handling hierarchically encoded data as well as the association of meta-information about the content (called 'object content information') and the intellectual property rights associated with it.

[0029] Each stream itself is characterized by a set of descriptors for configuration information, e.g., to determine the required decoder resources and the precision of encoded timing information. Furthermore the descriptors may carry hints to the Quality of Service (QoS) it requests for transmission; e.g., maximum bit rate, bit error rate, priority, etc.

[0030] Synchronization of elementary streams is achieved through time stamping of individual access units within elementary streams. The synchronization layer manages the identification of such access units and the time stamping. Independent of the media type, this layer allows identification of the type of access unit; e.g., video or audio frames, scene description commands in elementary streams, recovery of the media object's or scene description's time base, and it enables synchronization among them. The syntax of this layer is configurable in a large number of ways, allowing use in a broad spectrum of systems.

Delivery of streaming data

[0031] The synchronized delivery of streaming information from source to destination, exploiting different QoS as available from the network, is specified in terms of the synchronization layer and a delivery layer containing a two-layer multiplexer.

[0032] The first multiplexing layer is managed according to the DMIF specification. (DMIF stands for Delivery Multimedia Integration Framework) This multiplex may be embodied by the MPEG-defined FlexMux tool, which allows grouping of Elementary Streams (ESs) with a low multiplexing overhead. Multiplexing at this layer may be used, for example, to group ES with similar QoS requirements, reduce the number of network connections or the

end to end delay.

[0033] The "TransMux" (Transport Multiplexing) layer offers transport services matching the requested QoS. Only the interface to this layer is specified by MPEG-4 while the concrete mapping of the data packets and control signaling must be done in collaboration with the bodies that have jurisdiction over the respective transport protocol. Any suitable existing transport protocol stack such as (RTP)/UDP/IP, (AAL5)/ATM, or MPEG-2's Transport Stream over a suitable link layer may become a specific TransMux instance. It is possible to:

- Identify access units, transport timestamps and clock reference information and identify data loss.
- Optionally interleave data from different elementary streams into FlexMux streams
- Convey control information to:

- indicate the required QoS for each elementary stream and FlexMux stream;
- translate such QoS requirements into actual network resources;
- associate elementary streams to media objects
- Convey the mapping of elementary streams to FlexMux and TransMux channels

Interaction with media objects

[0034] In general, the user observes a scene that is composed following the design of the scene's author. Depending on the degree of freedom allowed by the author, however, the user has the possibility to interact with the scene. Operations a user may be allowed to perform include:

- change the viewing/listening point of the scene, e. g. by navigation through a scene;
- drag objects in the scene to a different position;
- trigger a cascade of events by selecting a specific object, e.g. starting or stopping a video stream;
- select the desired language when multiple language tracks are available;

[0035] The multimedia content delivery chain encompasses content creation, production, delivery and consumption. To support this, the content has to be identified, described, managed and protected. The transport and delivery of content will occur over a heterogeneous set of terminals and networks within which events will occur and require reporting. Such reporting will include reliable delivery, the management of personal data and preferences taking user privacy into account and the management of (financial) transactions.

[0036] The MPEG-21 multimedia framework identifies and define the key elements needed to support the multimedia delivery chain as described above, the relationships between and the operations supported by them. MPEG-21, MPEG will elaborate the elements by defining the syntax and semantics of their characteristics, such as interfaces to the elements. MPEG-21 will also address the necessary framework functionality, such as the protocols associated with the interfaces, and mechanisms to provide a repository, composition, conformance, etc.

[0037] The seven key elements defined in MPEG-21 are:

- Digital Item Declaration (a uniform and flexible abstraction and interoperable schema for declaring Digital Items);
- Digital Item Identification and Description (a framework for identification and description of any entity regardless of its nature, type or granularity);
- Content Handling and Usage (provide interfaces and protocols that enable creation, manipulation, search, access, storage, delivery, and (re)use of content across the content distribution and consumption value chain);
- Intellectual Property Management and Protection (the means to enable content to be persistently and reliably managed and protected across a wide range of networks and devices);
- Terminals and Networks (the ability to provide interoperable and transparent access to content across networks and terminals);
- Content Representation (how the media resources are represented);
- Event Reporting (the metrics and interfaces that enable Users to understand precisely the performance of all reportable events within the framework).

Problem

[0038] Content and service providers as well as end users demand for remote provisioning (at the providers facilities) of high-quality entertainment services. State-of-the-art video gaming and future virtual reality based applications will generate requirements on high-dynamic, interactive, and high-resolution audio/video. Real-time video/audio processing for electronic gaming and other interactive virtual reality based entertainment requires specialized and performant local resources, e.g. PCs or game consoles.

[0039] The problem to be solved by the invention is the provisioning of highly interactive video/audio services for end users, e.g. remote gaming, with reactive requirements and hard real-time conditions. Challenging is the real-time behavior on user commands and a required reactive and a realistic dynamic visualization.

[0040] The solution should embed in the existing environment. I. e. remote hosted service, e.g. video games, should be based on the standard broadcast TV distribution concepts and therefore designed additional control path for the user interaction like MHP.

[0041] Currently there are no adequate solutions for individual interactive virtual reality services, because

the response time seems not allow realistic dynamic behavior, and the exhaustive motion in the video stream exhausting bandwidth.

[0042] Remote hosted simple video games based on the standard broadcast TV distribution path and an additional control path are known, but they provide no adequate solutions for individual interactive services, because the response time does not allow realistic dynamic behavior.

BRIEF DESCRIPTION OF THE INVENTION

[0043] The invention provides an Action Streaming Environment for end-users. That is an interactive streaming service and an interaction device enabling the user to interact in a virtual reality. Simple interaction devices, e.g. a set-top-box, are used to subscribe and participating on a personalized interactive streaming service which is provided on a centralized server accessible via a broadband access network.

[0044] Action services are individually and interactive, in real time composed audio/video streams, e.g. direct interaction with an avatar, including multi-user virtual environments or realities, e.g. for online gaming or virtual cities in a realistic animation allowing direct user/user and user/environment interaction within the environment.

Hardware Prerequisite

[0045] An end-user needs a set-top-box or TV integrated digital audio/video signal handling facilities for receiving TV broadcast and individual channels, e.g. for video on demand. The format used of the remote generated broadband entertainment stream should be compatible with the available digital audio/video signal handling facilities, e.g. MPEG, DVB-MHP compliance.

Functional Requirements

[0046] The end-user interaction requires a control channel in the reverse direction. The end-user's equipment sends the stimuli or commands to the entertainment service specific processing elements. Enhancements of the user equipment may be realized by downloading the new functionality including session oriented function downloads driven by the service environment.

[0047] The action streaming service will be originated at the remote (central) location by service specific processing elements for generating the audio/video streams for multiple end-users.

Network Requirements

[0048] For the individual downstream channel towards the user, guaranties according to bandwidth and service delivery time are required for the operation. The individual control path in the reverse direction primarily must meet especially the delay constraints, in order to keep the user-service-user response time under the perceptible limits. It is essential controlling the access network elements according to the required quality of service parameters. I.e., the service environment generally and/or on a session specific bases has to request the setup of the data paths with the required service quality level at the access network control entities liable for the media stream transport.

OBJECTS AND ADVANTAGES OF THE INVENTION

[0049] The invention is a **Method** for generating an interactive virtual reality with real time user interaction for at least one individual user client at a network service using interactive media-streaming technology comprising the steps of establishing an action stream session comprising connection handling between said network service and said client, quality of service handling, adapting the network environment by demanding network resources and control information in the user's client and participated network elements, establishing media-streaming path from the service to the client and a user interaction control path in the reverse direction, generating and transmitting individual media streams to the client by embedding interaction into a virtual reality, and extracting and encoding a media stream at the service using a virtual reality description compressed motion picture stream, encoding and transmitting the user's interaction to the service, as well as de-coding and playing the individual media data stream at the client side.

[0050] The network environment and the media streaming path might be coordinated between for multiple possibly interacting user clients. The network might be controlled for ensuring the required quality of service and possibly interacting user clients and based on the virtual reality scenario might be coordinated. The quality of service might be especially high data rates in the downstream direction as well as a minimal round trip delay in both directions. This requires a delay minimized encoding of the media stream, e.g. on a frame by frame basis, even for compressed media formats.

[0051] The generating of individual media streams by embedding interaction into a virtual reality, and extracting and encoding a media stream at the service using a virtual reality description compressed motion picture stream might be performed by coding parts of the virtual reality description, e.g. as requested by a game application by a hardware independent audio-visual programming interface like Microsoft's DirectX, directly in the outgoing compressed data stream. The media stream might be on the basis of an application oriented graphic and/or sound description information without intermediate uncompressed video information. The action stream session might comprising an compatibility alignment, e.g. by updating and configuring software parts of the service and/or the client by uploading necessary software.

[0052] The invention is an **Action Streaming Client** for generating an interactive virtual reality with real time user interaction using interactive media-streaming technology comprising a downstream interface for receiving interactive media streams, decoding means for viewing interactive media streams, and controlling means for encoding user interaction and demanding network resources, a upstream interface for transmitting the encoded interaction leading to an instantaneous manipulation of the media downstream channel.

[0053] The Action Streaming Client might be realized by a device with DSL access enabled digital TV user equipment or it might be realized by a enabled personal computer with DSL access.

[0054] The invention is an **Action Streaming Server** for generating an interactive virtual reality with real time user interaction using interactive media-streaming technology comprising at least one upstream interface for receiving user interaction and at least one downstream interface for providing an interactive media-stream, and for each user an interpreter for the received user interaction, a virtual reality engine for embedding the user interaction in the virtual reality, a media extraction part for extracting an individual media stream an encoder for encoding the individual media stream, and (commonly shared) a network controlling unit for ensuring the required quality of service, continuously, and a environment controller for coordinating multiple individual virtual realities, and multiple individual media streams.

[0055] The invention is an **Action Streaming Service** providing resources for generating an interactive virtual reality with real time user interaction using interactive media-streaming technology comprising at least one upstream interface for receiving user interaction and at least one downstream interface for providing an interactive media-stream, and for each user an interpreter for the received user interaction, a virtual reality engine for embedding the user interaction in the virtual reality, a media extraction part for extracting an individual media stream, an encoder for encoding the individual media stream, and (commonly shared) a network controlling unit for ensuring the required quality of service, continuously, and a environment controller for coordinating multiple individual virtual realities, and multiple individual media streams.

[0056] The invention is a **Action Streaming System** for generating an interactive virtual reality with real time user interaction using interactive media-streaming technology comprising an access network providing at least one action streaming service, where said action streaming service comprises means for generating an interactive virtual reality, at least one action streaming client comprises means for consuming said interactive virtual reality where said service is located in a network at a action streaming server and said network is controlled said the service.

[0057] The invention is an **Action Stream** comprising a data structure for encoding, decoding a virtual reality in a media data stream a data structure for embedding interaction, and a control structure for managing network resources ensuring the required quality of service.

[0058] The Action Stream might be realized by a Digital Video Broadcast Multimedia Home Platform compliant video/audio and control data stream. The Action Stream might be realized MPEG compliant video/audio and control data stream.

[0059] The invention is an **Action Streaming Session** comprising a connection handling between service and client, a perquisite quality of service handling ensuring that the network provides the required quality of service, and a continuos quality of service handling according to the service's quality of service demands, a compatibility alignment between server and client, a service authentication-authorization-and-accounting, as well as action stream exchange.

[0060] The invention is an **Action Streaming Protocol** comprising means for creating an action streaming service session, means for adaptation of the users' client and the service, means for authentication-authorization-and-accounting, means for controlling network resources according to quality of services demands, and means for coordinating and exchanging action streams.

[0061] And the invention are **Computer Software Products** for generating an interactive virtual reality with real time user interaction using interactive media-streaming technology realizing an Action Streaming Service and an Action Streaming Client.

[0062] Accordingly, it is an object and advantage of the present invention to provide novel interactive services for subscribers: gaming, information services, remote-learning, etc. based on emerging virtual reality/worlds technology, i.e. a user-controlled real time composed video stream.

[0063] Another advantage of the present invention is that only few equipment at the subscriber site is required in addition to prevalent MPEG aware TV equipment. Especially no need for an expensive video game console and a broad spectrum of pay-per-use games.

[0064] A further advantage of the present invention is that it is Multimedia Home Platform (DVB-MHP) compliant. Broadband entertainment is expected to be the future of business of service providers. And the inventions uses the broadband infrastructure enabling a shared resource, i.e. an action service with the relatively low cost of an individual broadband access.

[0065] These and many other objects and advantages of the present invention will become apparent to those of ordinary skill in the art from a consideration of the drawings and ensuing description.

BRIEF DESCRIPTION OF THE FIGURES

[0066]

Figure. 1 illustrates a prior art combination of the

three main types of picture decomposition used in MPEG-1.

Figure. 2 shows a prior art MPEG scene description builds on several concepts from the Virtual Reality Modeling Language in terms of both its structure and the functionality of object composition nodes.

Figure. 3 is a schematic drawing of the networking context of an action streaming environment with the components according to the invention.

Figure. 4 is a drawing of an action streaming server according to the invention.

Figure. 5 is a schematic drawing of the architecture of the action streaming server according to the invention.

Figure. 6 is a drawing of an action streaming client according to the invention.

Figure. 7 is a schematic drawing of the architecture of the action streaming client according to the invention.

DETAILED DESCRIPTION OF THE INVENTION

[0067] Those of ordinary skill in the art will realize that the following description of the present invention is illustrative only and is not intended to be in any way limiting. Other embodiments of the invention will readily suggest themselves to such skilled persons from an examination of the within disclosure.

[0068] **Figure. 1** shows a sequence SEQ of pictures with a consecutive sub-sequence or group GRP of pictures. It shows a single picture PIC comprising a horizontal slice SLC, consisting of blocks. It further shows a macroblock MBC consisting of multiple blocks and a single block BLC.

[0069] The drawing illustrates main types of picture decomposition used in MPEG-1. In a continuous picture sequence SEQ only the varying parts carry information. To extract and identify these parts a picture sequence SEQ is decomposed into groups GRP and a picture PIC is decomposed into slices SLC, macroblocks MBC and blocks BLC. This fact is heavily used to save network and memory resources when transmitting or storing video data.

[0070] **Figure. 2** shows a prior art MPEG scene description builds on several concepts from the Virtual Reality Modeling Language in terms of both its structure and the functionality of object composition nodes. The drawing contains a virtual reality consisting of 2- and 3-dimensional audio visual objects OBJ originated and controlled by a multiplexed downstream DS and streamed into an encoded multiplexed upstream US. The scene comprises a coordinate system CS, and the audio visual objects OBJ in the space generated by the scene coordinate system CS are projected onto a projection plane PP with respect to a hypothetical viewer VW. Video information VI is extracted with respect to this projection and audio information AU is extracted correspondingly by integrating the audio objects into a so called psycho acoustic model.

[0071] The drawing illustrates how a virtual reality consisting of audio visual objects can be manipulated object-wise by streamed control data DS, how these objects origin streamed control data US, and how audio streams AU and video streams VI could be derived. Note that the object-wise presentation of the virtual reality is natural and enables a tight encoding.

[0072] **Figure. 3** shows a schematic drawing of the networking context of the invention. It contains a network NW consisting of network access points AP, e.g. an network access server and network elements NE, e. g. switches, routers, gateways, etc. Furthermore the network comprises an action streaming service system provided by an action streaming server ASS. The network elements including the action streaming server and the network access points are inter-connected by means of network connection paths or channels CN, illustrated by the arrows. The network access points APs provide action streaming clients ASCs access to the network NW via an access line AC, e.g. digital subscriber line (DSL), illustrated by thin arrows.

[0073] The thick end of the arrows depicting the channels CN modeling a (broadband) downstream carrying media information of a virtual world generated by the action streaming service, and the thin ends modeling the upstream carrying user interactions originated by the action streaming clients ASC. The action streaming service ASS controls the network elements by demanding necessary quality of services and indirectly defining connection paths CN ensuring a high-quality interactive virtual reality at the action streaming clients. Downstream and upstream data may be routed on different paths to and from an ASC. Network controlling connections might be routed on different paths, too.

[0074] **Figure. 4** is a drawing of an action streaming server ASS. It shows a network channel interfaces IN the network NW (environment), as well as a drawing of a computer cluster providing the virtual realities and the corresponding video and audio streams for a plurality of action streaming clients (generic supercomputer for visualization applications or computer blades based on game console technology).

[0075] **Figure. 5** is a schematic drawing of the architecture of an action streaming service system provided on an action streaming server. It shows a service environment and network controller ENV_CON and a plurality of session controller SES-CON managing a quadruple of units, a stimuli injection unit INJ, a virtual reality engine VRE, a media extraction unit ME, and a video streaming encoding unit VSE.

[0076] The service environment and network control-

ler ENV_CON controls the service environment comprising of coordinating multiple, maybe one common shared, virtual realities. It controls the session interaction comprising all shown units INJ, VRE, ME, VSE with respect to performed actions. Multi-player environments can be implemented either tight or lose coupled, i.e. all users sharing the same session or a session per user coupled by inter-session communication. It has to take into account the desirable, granted, and available quality of services or network resources using e.g., common video streams and broadcasting or balancing the load of the single session processors. It could be even advantageous to allocate certain processing to an action streaming client. Such a concept is closely linked with the audio-visual coding standard used, e.g. in contrast to the mainly video oriented MPEG-1 and -2 standards MPEG-4 offers flexible combinable media objects.

[0077] The session controller SES-CON is responsible, e.g. for authentication-authorization-and-accounting tasks, for connection setup, for the choice of the virtual reality, for the client-service alignment, etc. It is the controlling instance of the provided action stream.

[0078] The action stream is generated by the four logical processing units, named stimuli injection INJ, virtual reality engine VRE, media extraction ME, and video stream encoding. The stimuli injection INJ receives the users interaction from the network and translates it for the virtual reality engine. The virtual reality engine VRE creates continuously new states based on the state history and the stimuli injections. This timed scene is called virtual reality or virtual world. It can consist of buildings, tools, backgrounds, a garden, streets, a play ground, a star ship, a compass, or any audio visual object. It even can comprising the user simulated itself. It can provide state information and feedback, e.g. force feedback for a joystick, a visual impression, e.g. a video, sound, or any reactivity in general. The view for the subscribed action streaming client is extracted from virtual reality model by the media extraction ME.

[0079] And it is encoded into media/command stream by the video stream encoding VSE. The drawing illustrates for simplicity reasons only the video encoding although all media could be encoded analogously.

[0080] The action streaming service system ASS might comprising hardware implemented algorithms for the direct generation of the compressed media stream from an application oriented graphic and/or sound description information. Thus avoiding an intermediate uncompressed video information as generated by ordinary visual processors (3D graphics accelerators).

[0081] **Figure. 6** is a drawing of an action streaming client ASC comprising several input devices, here a joystick JS a remote control RC a keyboard KB and a joypad JP. The client itself is realized by a set-top box STB. The set-top box has a interface connection AC to a network access point, e.g., a digital subscriber line, providing access to the network NW.

[0082] The drawing illustrates the idea of a very simple and cheap (compared with a complex expensive high tech game console like a play station or a game cube) customer premises equipment, reusing a television set. The action streaming client realizing device implementing a DSL access enabled digital TV user equipment might be integrated in a next generation TV set instead of being a set-top box.

[0083] Alternatively customers using standard PC equipment can get on-demand access to the whole set of recent games, without the need to invest permanently in the top end of graphics accelerators and CPU technology.

[0084] **Figure. 7** is a schematic drawing of the architecture of the action streaming client ASC. The action streaming client ASC comprising a transport protocol and physical interface TP/PI to the network NW, it comprises a plurality of media players ME-P and a graphics unit GR. It comprises a user interaction unit UI, managing remote control RC keyboard KB joypad JP, joystick JS, etc. input IN. The graphics and the media player provide video VI, audio AU, etc. output OU. The media player are coordinated by a media control unit ME-CT for aligning and synchronizing the multi-media. Further, the architecture comprises an information and data access IDA. In the center of these components a application APP is embedded using, instructing and coordinating said components.

[0085] Operationally the action streaming client receives from the network NW media streams using a physical interface PI and a transport protocol TP, commands for the running action streaming application APP is provided via information and data access. The Application might coordinate the media streams via the media control ME-CT. User interaction from human user interface devices are provided to the application APP via the user interaction component UI. This architecture is similar to the multi media home platform.

[0086] Future applications for end users a novel class of highly interactive virtual reality applications. Action services are individually and interactive, in real time composed media streams, e.g. direct interaction with an avatar including multi-user virtual environments, e.g., for online gaming or virtual cities in a realistic animation allowing direct user/user and user/machine interaction.

[0087] An end-user needs no complex equipment with a high technology drive, only a set-top-box or TV integrated digital audio/video signal handling facilities for receiving TV broadcast channels. The format used for transmitting the remote generated broadband entertainment stream should be compatible with the available digital audio/video signal handling facilities, e.g. using a MPEG (Motion Pictures Expert Group) family standard format compliant with the Multimedia Home Platform. The end-user interaction requires a channel in the reverse direction. The end-user equipment converts the stimuli / commands derived from human interfaces to an action stream service control protocol data flow. Enhancements of the user equipment may be realized by

downloading the new functionality including session oriented function downloads driven by the service environment.

[0088] The action streaming service will be originated at a remote (central) location by service specific processing elements for generating the media streams for multiple end-users. The information derived from the media processing function block has to be converted / encoded by adaptation means into the downstream digital media signal as required by the user equipment. This might be done on a delay minimized frame by frame basis, even for compressed video formats. An efficient way generating the output stream seems the direct translation of the description format for the audiovisual effects as used by the application defined within the service environment / operating system as application programming interface into the coding of the media stream.

[0089] Input for the (inter-)action streaming respectively entertainment service generation is the service control protocol relaying the user stimuli. Adaptation here means terminating the control protocol and emulating local input / steering means support the porting of, e.g. gaming, applications designed for the local usage.

[0090] The action streaming technology makes high demands on the access network between user and the location of the service origination. For the individual downstream channel towards the user, guaranties according to bandwidth and service delivery time are required for the operation. The individual control path in the reverse direction primarily must meet especially the delay constraints, in order to keep the user-service-user response time under the perceptible limits. Complying these network related quality of service parameters is advantageous for the service quality and finally the service acceptance. The access network elements realizing the data paths must be controlled according to the required quality of service parameters. Action streaming quality of service requirements have to be requested/ controlled (generally and/or user session specific) by the service environment, e.g. using an access network data path control.

Alternative Embodiments

[0091] Although illustrative presently preferred embodiments and applications of this invention are shown and described herein, many variations and modifications are possible which remain within the concept, scope, and spirit of the invention, and these variations would become clear to those of skill in the art after perusal of this application.

[0092] Alternatively, the invention can be used with any type of media and enabled action streaming client. In future it is expectable to have devices stimulating more senses more perfect, e.g. having hologram projectors, aura generators, wearable suits providing impressions like temperature , pressure, or vibration to a user's sense of touch.

[0093] The invention, therefore, is not intended to be limited to audio or video except in the spirit of the appended claims.

Claims

1. A **Method** for generating interactive individual virtual reality with at least one action streaming client (ASC) and a network action streaming service system using interactive media-streaming technology comprising the steps of
   - establishing an action stream session comprising connection handling between said network service and said client, quality of service handling,
   - adapting the network environment by demanding network resources and control information in the user's client and participated network elements (NE),
   - establishing media-streaming path (CN) from the service to the client and a user interaction control path (CN) from the client to the service,
   - generating and transmitting individual media streams to the client by embedding interaction into a virtual reality, and extracting and encoding a media stream at the service using a virtual reality description compressed motion picture stream,
   - encoding and transmitting the user's interaction to the service, as well as de-coding and playing the individual media data stream at the client side.

2. The **Method** according to Claim 1, where said network environment and said media streaming path is coordinated for multiple possibly interacting user action streaming clients (ASC).

3. The **Method** according to Claim 1, further comprising the step of controlling the network for controlling the required quality of service, continuously.

4. The **Method** according to Claim 3, where said controlling the network for ensuring the required quality of service is coordinated between for multiple possibly interacting user action streaming clients (ASC) and based on the virtual reality scenario.

5. The **Method** according to Claim 3, where said quality of service are especially high data rates in the downstream direction as well as a minimal round trip delay and/or minimal delay variation in both directions.

6. The **Method** according to Claim 1, where said gen-

erating individual media streams by embedding interaction into a virtual reality, and extracting and encoding a media stream at the service using a virtual reality description compressed motion picture stream is performed by coding parts of the virtual reality description directly in the outgoing compressed data stream.

7. The **Method** according to Claim 1, where said media stream is on the basis of an application oriented graphic and/or sound description information without intermediate uncompressed video information.

8. The **Method** according to Claim 1, where said action stream session comprising an compatibility alignment, e.g. by updating and configuring software parts of the service and/or the client by uploading software.

9. An **Action Streaming Client** (ASC) for generating interactive individual virtual reality invoking a network action streaming service system using interactive media-streaming technology comprising

- a downstream interface (TP/PI) for receiving interactive media streams,
- decoding means for viewing interactive media streams, and
- controlling means (APP, ME-CT, ME-P, TP/PI, UI) for encoding user interaction and demanding network resources,
- a upstream interface (TP/PI) for transmitting the encoded interaction, leading to an instantaneous manipulation of the media downstream channel.

10. The **Action Streaming Client** (ASC) according to Claim 9, realized by a device implementing a DSL access enabled digital TV user equipment.

11. The **Action Streaming Client** (ASC) according to Claim 9, realized by a enabled personal computer with DSL access.

12. An **Action Streaming Server** (ASS) for generating interactive individual virtual reality providing a network action streaming service system using interactive media-streaming technology comprising

- at least one upstream interface (IN) for receiving user interaction and at least one downstream interface (IN) for providing an interactive media-stream, and
- for at least one user sharing a media stream

  - an interpreter (INJ) for the received user interaction,
  - a virtual reality engine (VRE) for embedding the user interaction in the virtual reality,
  - a media extraction (ME) part for extracting an individual media stream,
  - an encoder (ME) for encoding the individual media stream, and

- a session controlling unit (SES-CON) for controlling the required quality of service, continuously,
- and a environment controller (ENV-CON) for coordinating multiple individual virtual realities, and multiple individual media streams.

13. An **Action Streaming Service System** (ASS) providing resources for generating interactive individual virtual reality using interactive media-streaming technology comprising

- at least one upstream interface (IN) for receiving user interaction and at least one downstream interface (IN) for providing an interactive media-stream, and
- for at least one user sharing a media stream

  - an interpreter (INJ) for the received user interaction,
  - a virtual reality engine (VRE) for embedding the user interaction in the virtual reality,
  - a media extraction (ME) part for extracting an individual media stream,
  - an encoder (VSE) for encoding the individual media stream, and

- a session controlling unit (SES-CON) for ensuring the required quality of service, continuously,
- and an environment controller (ENV-CON) for coordinating multiple individual virtual realities, and multiple individual media streams.

14. An **Action Streaming System** for generating an interactive virtual reality with real time user interaction using interactive media-streaming technology comprising

- an access network (NW) providing at least one action streaming service system,
- where said action streaming service system (ASS) comprises means for generating an interactive virtual reality,
- at least one action streaming client (ASC) comprises means for consuming said interactive virtual reality,
- where said action streaming service is located in a network at an action streaming server (ASS) and said network (NW) is controlled by the service.

15. An **Action Stream** comprising a data structure for encoding, decoding a virtual reality in a media data stream a data structure for embedding interaction, and a control structure for managing network resources ensuring the required quality of service.

16. The **Action Stream** realized by a Digital Video Broadcast Multimedia Home Platform compliant video/audio and control data stream.

17. The **Action Stream** realized MPEG compliant video/audio and control data stream.

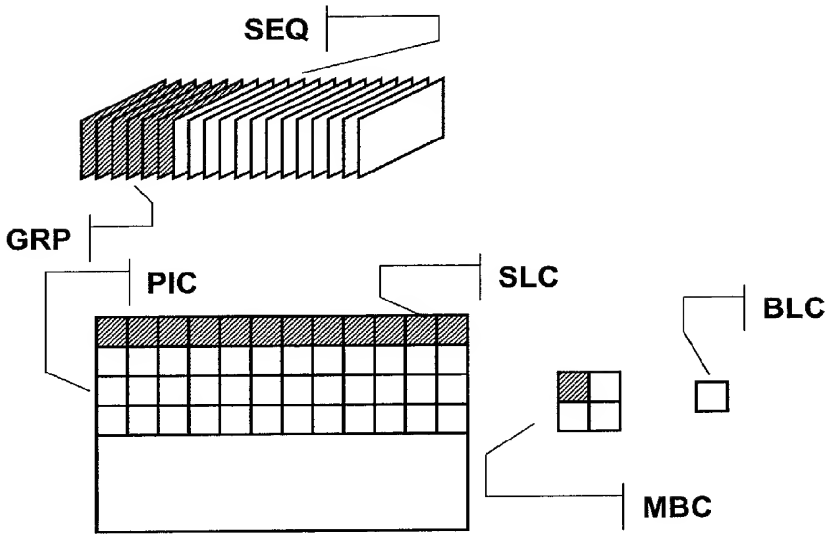
18. An **Action Streaming Session** comprising

- a connection handling between service and at least one client
- a perquisite quality of service handling ensuring that the network provides the required quality of service, and
- a continuos quality of service handling according to the service's quality of service demands,
- a compatibility alignment between server and client,
- a service authentication-authorization-and-accounting,
- as well as action stream exchange.

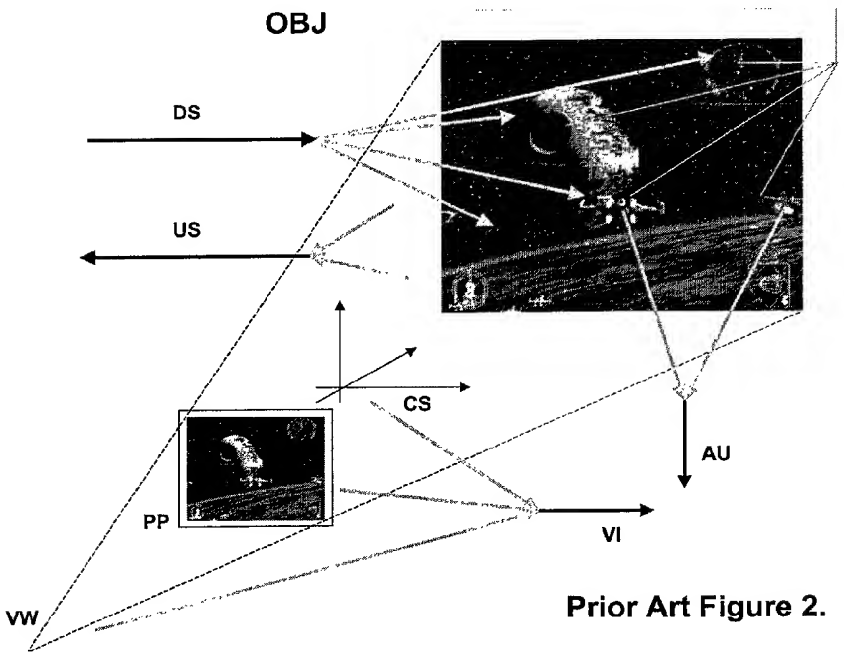
19. An **Action Streaming** Protocol comprising means for creating an action streaming service session, means for adaptation of the users' client and the service, means for authentication-authorization-and-accounting, means for controlling network resources according to quality of services demands, and means for coordinating and exchanging action streams.

20. A **Computer Software Product** for generating an interactive virtual reality with real time user interaction using interactive media-streaming technology realizing an Action Streaming Service according to claim 13.

21. A **Computer Software Product** for generating an interactive virtual reality with real time user interaction using interactive media-streaming technology realizing an Action Streaming Client according to claim 9.



Prior Art Figure 1.



Prior Art Figure 2.

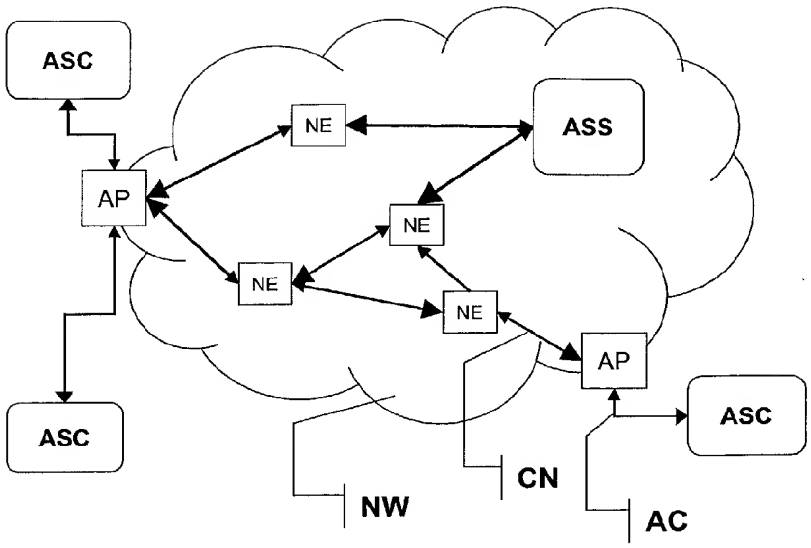


Figure 3.

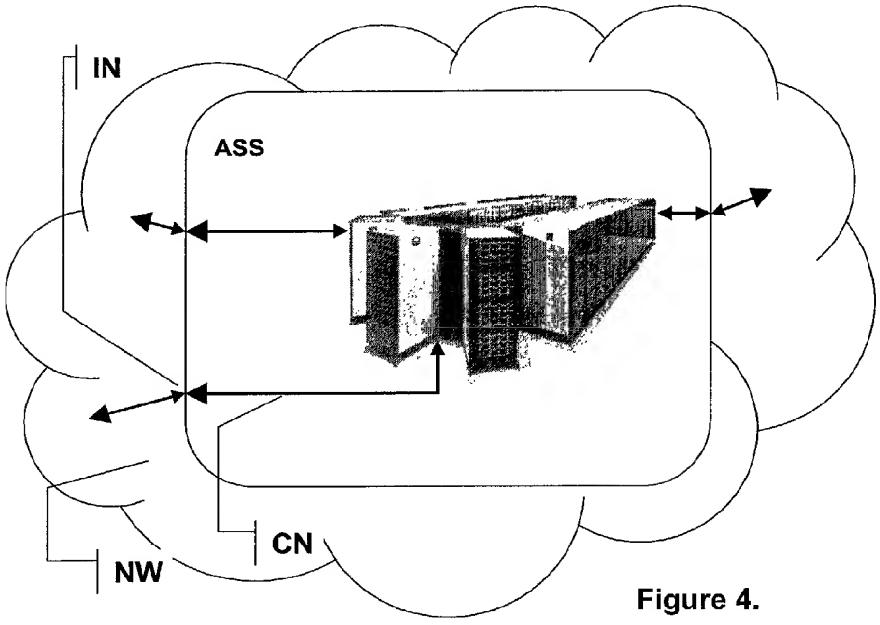


Figure 4.

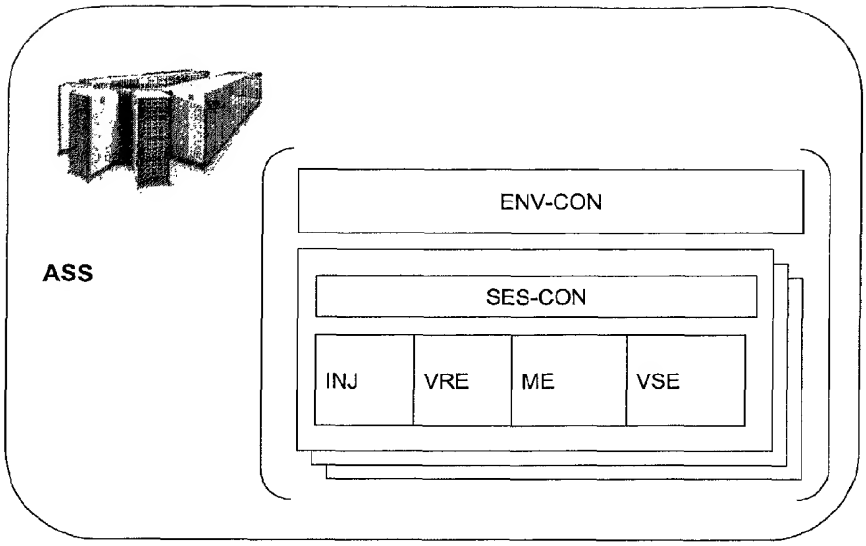


Figure 5.

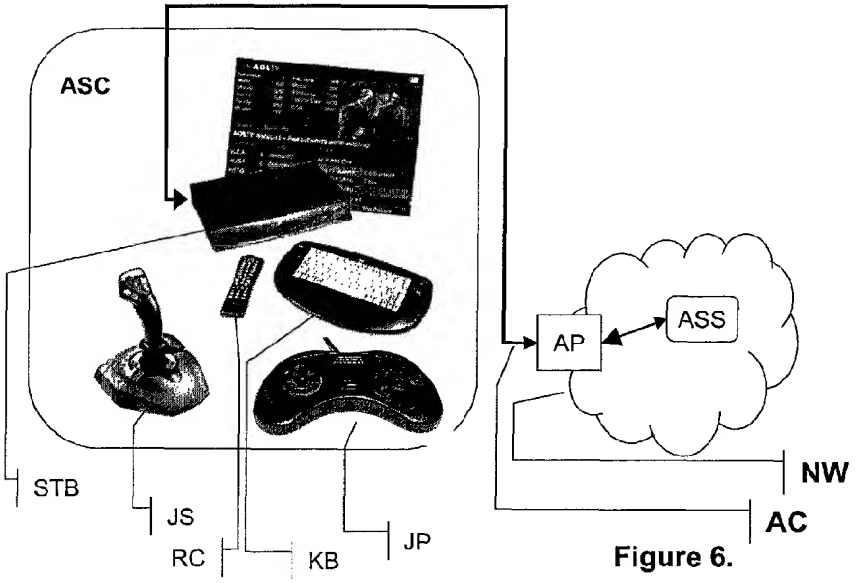


Figure 6.

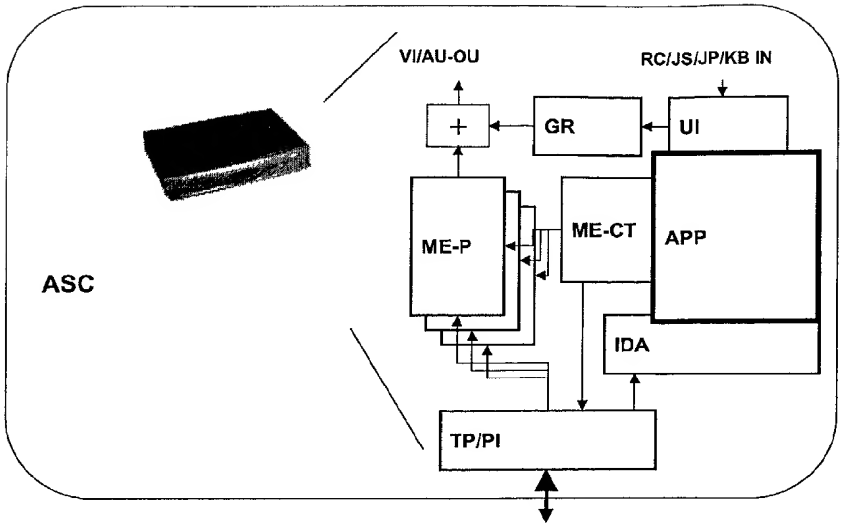


Figure 7.

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 02 36 0239

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.7)
X	DOENGES P K ET AL: "Audio/video and synthetic graphics/audio for mixed media" SIGNAL PROCESSING. IMAGE COMMUNICATION, ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, NL, vol. 9, no. 4, 1 May 1997 (1997-05-01), pages 433-463, XP004075338 ISSN: 0923-5965	1,2,6,7, 9,12-15, 17,20,21	A63F13/12 H04N7/24 H04N7/173
Y	* page 436, left-hand column, paragraph 2 * * page 437, left-hand column, paragraph 1 * * page 438, left-hand column, paragraph 2.3 * * page 441 - page 442, paragraph 2.6 * * page 444, left-hand column, paragraph 2 * * page 444, right-hand column, paragraph 2 - page 445, left-hand column, paragraph 1 * * page 456, right-hand column, paragraph 6 * * page 458, left-hand column, paragraph 6.4 - page 460, right-hand column, paragraph 6.5 * * figure 4 * ---	3,18,19	
Y	HUARD J-F ET AL: "REALIZING THE MPEG-4 MULTIMEDIA DELIVERY FRAMEWORK" IEEE NETWORK, IEEE INC. NEW YORK, US, vol. 12, no. 6, November 1998 (1998-11), pages 35-45, XP000873126 ISSN: 0890-8044 * page 37, left-hand column, paragraph 6 - right-hand column, paragraph 1 * --- -/--	3,18,19	

TECHNICAL FIELDS SEARCHED (Int.Cl.7)
A63F
H04N

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	17 January 2003	Sindic, G

CATEGORY OF CITED DOCUMENTS
X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 02 36 0239

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.7)
A	"Multimedia Home Platform" [ONLINE], November 2001 (2001-11), pages 1-12, XP002224935 * page 3, paragraph 6 * ---	10,11,16	
A	WO 98 57718 A (CORNWELL SIMON ANTHONY VIVIAN ;KYDD RICHARD ANDREW (GB); WRIGHT DA) 23 December 1998 (1998-12-23) * page 13, line 22 - line 37 * * page 15, line 35 - page 16, line 1 * * page 16, line 15 - page 17, line 6 * ---	1,6,7,9, 12-19	
A	WO 00 11847 A (KONINKL PHILIPS ELECTRONICS NV) 2 March 2000 (2000-03-02) * page 2, line 16 - line 30 * * page 8, line 7 - line 22 * -----	1,5,9, 12-19	
			TECHNICAL FIELDS SEARCHED (Int.Cl.7)

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	17 January 2003	Sindic, G

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

ANNEX TO THE EUROPEAN SEARCH REPORT ON EUROPEAN PATENT APPLICATION NO. EP 02 36 0239

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report. The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

17-01-2003

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9857718 A	23-12-1998	AT 216909 T	15-05-2002
		AU 729838 B2	08-02-2001
		AU 8118098 A	04-01-1999
		DE 69805176 D1	06-06-2002
		DE 69805176 T2	07-11-2002
		EP 0989892 A1	05-04-2000
		WO 9857718 A1	23-12-1998
		JP 2002508095 T	12-03-2002
		NZ 501348 A	28-04-2000
		PT 989892 T	31-10-2002
WO 0011847 A	02-03-2000	BR 9906766 A	03-10-2000
		CA 2306785 A1	02-03-2000
		CN 1275091 T	29-11-2000
		CN 1277774 T	20-12-2000
		WO 0011847 A1	02-03-2000
		WO 0010663 A1	02-03-2000
		EP 1048159 A1	02-11-2000
		EP 1047481 A1	02-11-2000
		JP 2002523156 T	30-07-2002
		JP 2002523980 T	30-07-2002
		TR 200001074 T1	21-11-2000

EPO FORM P0459

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82